REMARKS

Rejection under 35 USC 102(b)

Claims 1, 4 to 9 and 12 to 20 were rejected under 35 USC 102(b) as being anticipated by an Donigata et al article entitled "D Blue: An advanced Enterprise Information, Search and Delivery System:, and also on the basis of public use or sale of the D. Blue System described in the above Donigata et al article. (A copy of the article accompanying the rejection alleges that the article was published on 1/1/00.)

Applicants request reconsideration of a rejection under 35 USC 102(b) on the basis of a Donaganta et al article entitled "d Blue: An Advanced Enterprise Information Search and Delivery System" for the following reasons:

A. The following showing of why the submission of the affidavit required by the Examiner was not earlier presented in compliance with 37 CFR 116(e).

The reason why an affidavit did not accompany the timely submission of the corrected copy of the "d Blue" article is that the applicants' attorney did not think it was necessary to submit an affidavit under 37 CFR 1.132 with the corrected copy of the "d Blue" article. Section 1.132 provides for introduction of evidence "on a basis not otherwise provided for must be by way of an oath or declaration under this section" (emphasized). The corrected copy of the article speaks for itself as to the issue of the correct date of its publication. The "d Blue" article is the publication of an independent publisher. A copy of the article containing the change in the publication date can be obtained on the internet from the publisher's website (at least it was available there until 4/2/09). Applicants' attorney does not see how an affidavit of the inventor Dr. Moon Kim is required for further verification of an article in the independent publisher's website. Just as the Examiner relied on the independent publisher's internet website for a copy

of the article, applicants' attorney could rely on that website for the corrected copy. For this reason, applicants' considered that an affidavit by one of the inventors was unnecessary.

Further, contents of the flawed copy of the "d Blue" article, provided by the Examiner, are inconsistent with publication of the article on a New Year's Day 01/01/2000 at the start of a new millennium. As shown in Appendix A, containing two pages of the copy of the "d Blue" article relied on by the Examiner and the copy contains the following copyright notice:

"Published May 11, 2007 – Reads 19885"

Copyright © 2008 SYS – CON Media. All rights Reserved"

In addition, the list of references contained on the copy containing reference to a Park et al article dated 2001 and a Chu-Carrol et al article dated 2002. The 01/01/2000 date the Examiner relies on precedes the publish copyright notice dates and predates publication dates of references cited in the article. Therefore it is clear the 01/01/2000 relied by the Examiner in his rejection cannot be the correct publication date.

As opposed to the inconsistencies of the dates of the Park et al and Chu-Carrol articles with the first page publication date in the Examiner's provided copy of the "d Blue" article, the publication dates of the Park et al and Chu-Carrol et al articles are consistent with the 10/21/2002 publication date listed on the first page of the applicant provided article. Furthermore as shown in Appendix B containing two pages of the applicant provided copy of the modified "d Blue" article, the copyright notice in the applicant provided article contains a publishing date consistent with the one on the first page of the article. However, the copyright notice date is inconsistent with both listed publication dates. The copyright notice is as follows:

"Published October 21, 2002 – Reads 21978"

Copyright © 2008 SYS – Media, Inc. All rights reserved"

Appendix C is a reproduction of the hard copy of the "d Blue" article stating that it was published in October of 2002 which is consistent with both the dates of publication contained in the modified "d Blue" article. Appendix D contains two pages of the article dated 4/2/2009 that contains a copyright notice consistent with the publication dates listed in the articles contained in Appendices B and C. Therefore, it is respectfully submitted that the October 21, 2002 date is the correct publication date backed by data in Appendices B, C and D.

The provisions of the hard copy of the "d Blue" article was not considered necessary prior to final rejection. As pointed out above, applicants' attorney considered that no affidavit was required with the timely submission of the Examiner provided internet copy of the "d Blue" article identified in Appendix B. The pages of Appendix D were only considered appropriate for submission after the inconsistency of the copyright notice date in Appendix B was discovered in the preparation of this response.

B. No affidavit was needed

For the above reasons, applicants' attorneys' response without an affidavit or declaration still is considered correct. The affidavit is not required under 37 CFR 1.132 because, as pointed out above, the introduction of the modified article provides the basis for its own introduction. It is a publication of an independent publisher. It is available on the internet in the Websphere Journal website. It corrects obvious inconsistencies contained in the Examiner provided copy of the article from the same website. Therefore, it does not constitute introduction of evidence on a basis not otherwise provided for thereby requesting the need of an affidavit to support its introduction.

C. A showing under 37 CFR 116(c) is not required

If an affidavit was unnecessary, a showing under 37 CFR 116(c) is not now needed. Furthermore, if an affidavit had been required, it related back to the timely submission of the publisher modified document when no showing under section 116(c) was needed. For the above reasons, applicants' attorneys response without an affidavit was (and still is) considered correct.

D. The claimed invention is not disclosed in the "d Blue" article

In the present application, customers' unsuccessful search queries are located and then analyzed in a self enhancing search system to improve future search results. As shown in Figure 4 of the present application, this self-enhancing search system includes: a search system log analyzer 400, which periodically looks through the search system log 402 to uncover customers unsuccessful search queries (queries of customers that did not turn up a sufficient number of references or which resulted in customer complaints); a relevant document finder 406 which, based on enhanced query terms provided by a query analyzer 404, finds relevant documents 410 and 412 that were not found using the unsuccessful search queries; and a meta/data enhancer 408 that enhances the textual index for the relevant documents by adding to those relevant documents 410 and 412 terms (video player) used in the unsuccessful query to allow the relevant documents 410 and 412, turned up by the enhanced query, to be returned when future searches similar to, or the same as, the unsuccessful search queries are entered by users.

Figure 6 shows that along with search query terms (T(1,1), T(1,2) T(1,3),...,) that are found in each document (such as Doc #1) there are meta/data associated with each document that contains queries Q(1,1), Q(1,2), ... that are generated using the present invention and provided in the enhanced textual index. When a previously unsuccessful user query (say, Q(1,1)) is used to interrogate the database, the query Q(1,1)

interrogates both the search query terms found in each of the documents of the data base in step 702 and the meta/data search query terms for the documents in step 704 to identify relevant documents in steps 706 and 708. As a result, Doc #1 is identified as having meta/data containing the query Q(1,1). The results are then ranked in step 710 using not only original query words found in step 706, but also the modified query words obtained in step 708, and the results provided to the end user in step 712.

The applicants' attorney has reread the above-identified article and, contrary to the Examiner's position, nowhere did he find anything about the above described invention in the article. In fact, he did not find a mention of looking through the log analyzer for the purpose of locating failed or unsuccessful search queries for the purpose of enhancing the textual index of relevant documents not turned up by such failed or unsuccessful search queries. It does not describe the use of its log analyzer for locating failed search queries for the purpose of enhancing the textual index of the relevant documents not turned up by the failed search queries with search terms from the failed search queries so that later searches containing the failed queries will turn up the missed documents.

All independent claims in the application recite limitations that cover searching the search log of a database for unsatisfactory search queries and then adding search terms of such unsuccessful searches to applicable documents missed by the search. For instance, independent claims 1 and 9 call for: a search system analyzer system for looking through the search system log for unsuccessful customer queries; a relevant document finder for locating documents not found in the unsatisfactory search queries and the embedding of search terms of such unsuccessful search queries to documents missed by those unsuccessful queries but turned up by enhanced queries. Independent claim 17 calls for a search system analysis system for selecting unsuccessful customer search queries from a system log, a relevant document finder for identifying relevant documents not turned up by the unsuccessful search queries, and a meta/data

enhancer to link the relevant documents to search terms in unsuccessful searches that are not contained in the relevant documents so that when the original search terms are used in future queries these relevant documents will be found.

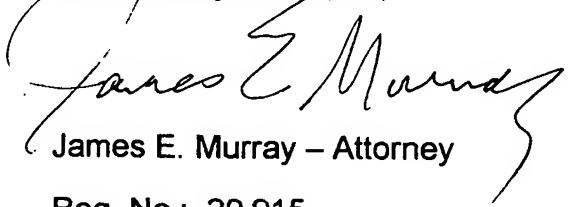
The dependent claims further distinguish over the description in the "d Blue" article adding details of the patentable limitations contained in the independent claims and add further limitations to the claimed subject matter.

In addition to not creating or disclosing a possible bar to filing under 35 USC 102(b), the article does not constitute a prior art reference that precludes patentability of the present invention under other sections of 35 USC 102. The inventors of the present invention are authors of the article, and the article does not disclose subject matter claimed in all the claims of the application.

For these and other reasons, the claims of this application are not barred by the contents of the Donigata et al article, and the existence of the article does not preclude their patentability under 35 USC 102 or 103.

For one, more or all of the above reasons, the Examiner is respectfully requested to reconsider the above-identified application, allow the application and pass it to issue.

Respectfully submitted,



James E. Murray – Attorney

Reg. No.: 20,915

Tel. No.: (845) 337-3199

APPENDIX A

Feature

dBlue - An Advanced Enterprise Information Search and Delivery System

The result of two years of work and five patentable inventions, dBlue is now available to IBM customers

By: Yurdaer Doganata; Lev Kozakov; Greg Brown; Tong-Haing Fin; Moon J. Kim; Youssef Drissi

(Article Publish Date: January 1, 2000) - One of the biggest complaints we hear about many company Web sites, from customers and employees alike, is that it's too hard to find what you need. At IBM, with 2.5 million Internet pages and more technical content than any single entity, including the Pentagon, that's no surprise.

A new IBM advanced information search and delivery system for the IBM support site (www.ibm.com/support) is expected to solve this problem. Code-named Digital Blue (dBlue), this project is a digital interface to IBM customers. The result of two years of work and five patentable inventions, dBlue is now available to IBM customers.

Remote Site Customization
Another dBlue feature that addresses corporate needs is Remote Site Customization (RSC). IBM, like any other large corporation, has multiple departments that may want to present search results and technical documents to their customers in different ways, adding their own ads, promotions, and so on. The dBlue system enables this by providing the RSC feature, which allows different departments to define their own layouts for search results and technical documents. The idea of RSC is rather simple: each remote site that wants to present the shared system content in a special format is allowed to store and register its own forms. When the system gets a request that specifies this remote site, it will use the appropriate form to build the customized view of the content. Figure 6 shows the six areas that are available for customization in a results page. To assist departments in customizing the layout of Web pages, dBlue provides a Web-based RSC administrative application, which allows the uploading and testing of customized forms.

Conclusion
dBlue has many advantages. In the near future, customers will be able to ask questions in natural language and the system won't require an exact match of words. In the near future, dBlue will also personalize searching so that once a user fills out a profile, responses will be filtered and ranked based on that profile. Multilanguage searches for documents written in Japanese, Chinese, and French will be supported by late 2002. By 2Q03, it's expected that all languages will be supported from a single Web application consistent with the vision of "one Web" for all regions.

Visit Author's Web Site: Yurdaer Doganata

References

- *IBM WebSphere Software Platform Overview: www7b. boulder.ibm.com/wsdd/products/platform overview.html*
- *WebSphere Everyplace Suite: www.ibm.com/pvc/prod ucts/ wes/index.shtml*
- *Booth, Alan E. Extending the Reach of Enterprise Applications with Transcoding and Machine Translation: www7b.software.ibm.com/wsdd/library/techarti cles/0206_booth/booth.html*
- *WebSphere Technology for Developers (overview & downloads): www7b.software.ibm.com/wsdd/downloads/ wstechnology_tech_preview.html*
- *Snell James. Implementing Web services with IBM WebSphere Version 4.0: www7b.software. ibm. com/wsdd/library/techarticles/0108_snell/0108_snell. html*
- *DB2 Product Family Overview: www.ibm.com/soft ware/data/db2/*
- *About the Internationalization Activity: www.w3.org/ International/about.html*
- *Internationalization in Java: http://java.sun.com/docs/ books/tutorial/i18n/*
- *The Unicode Home Page: www.unicode.org*
- *Park, Youngja; Byrd, Roy J.; and Boguraev, Branimir K. Automatic glossary extraction: Beyond terminology identification, IBM Research Technical Report RC22421, 2002. IBM T.J. Watson Research. The Talent (Text Analysis and Language Engineering) project: www.research.ibm.com/ talent*
- *Boguraev, Branimir K. and Neff, Mary S. (2000). Lexical Cohesion, Discourse Segmentation and Document Summarization. RIAO-2000. April.*
- *Park, Youngja and Byrd, Roy J. (2001). Hybrid text mining for finding terms and their abbreviations. EMNL-2001.*
- *Chu-Carrol, Jennifer; Prager, John; Ravin, Yael; and Cesar, Christian. (2002). A Hybrid Approach to Natural Language Web Services. EMNLP-2002.*

Published May. 11, 2007 — Reads 19,885

APPENDIX B

Feature

dBlue - An Advanced Enterprise Information Search and Delivery System

The result of two years of work and five patentable inventions, dBlue is now available to IBM customers

By: Yurdaer Doganata; Lev Kozakov; Greg Brown; Tong-Haing Fin; Moon J. Kim; Youssef Drissi

(Article Publish Date: October 21, 2002) - One of the biggest complaints we hear about many company Web sites, from customers and employees alike, is that it's too hard to find what you need. At IBM, with 2.5 million Internet pages and more technical content than any single entity, including the Pentagon, that's no surprise.

A new IBM advanced information search and delivery system for the IBM support site (www.ibm.com/support) is expected to solve this problem. Code-named Digital Blue (dBlue), this project is a digital interface to IBM customers. The result of two years of work and five patentable inventions, dBlue is now available to IBM customers.

The team that created dBlue is calling it "the next generation of enterprise information search-and-delivery systems." This is a WebSphere-based technology with breakthroughs in storing, searching, and retrieving information. Customers will be able to search for IBM technical support information using natural language and will receive results that are categorized, prioritized, and personalized. dBlue provides the foundation for a set of user-oriented support services applicable to all IBM support sites worldwide.

Rich Vazzana, vice president of ibm.com Support and Enablement, took on this project to improve the effectiveness and performance of IBM's Web-enabled post-sales support services. It became the underlying architecture of the "one-Web" vision across multiple IBM Web sites, improving adherence to IBM's company-wide standards and setting the stage for more advanced service offerings. The program will provide customers with IBM support experience, a single IBM support/service portal, toolset, and infrastructure. Hence, cross-IBM "common" support functions will be realized.

"The business goal is to improve goal achievement on the IBM Internet," said Frank Cummiskey, director of IBM eSupport & Services. "The primary reason that customers visit IBM's support sites is to resolve a technical problem. Today, only about 60% actually achieve their goal. Improving our customers' ability to find what they are looking for, as well as to find value in the information they find, will increase self-service on the Web, saving millions of dollars and increasing customer satisfaction."

System Architecture
Although dBlue architecture does not depend on the WebSphere software platform, it's the platform of choice of the dBlue architects for its

dynamically. Culturally dependent data, such as dates and currencies, appears in formats that conform to the end user's region and language.

The Unicode format, which handles most characters known to mankind, was instrumental in allowing the use of a unique globalized repository where multilingual searchable text and documents are encoded in one unique format. Unicode was also adopted as a standard format for encoding internal textual data in dBlue.

Localization
Localization (sometimes abbreviated as i10n) is the process of adapting software for a specific region or language by adding locale-specific components and translating text. Usually the most time-consuming part of the localization process is the translation of text. Other types of data, such as sounds and images, may require localization if they are culturally sensitive. Localizers also verify that the formatting of dates, numbers, and currencies conforms to local requirements.

Two innovative approaches in the globalization process are worth mentioning. The first allows documents to be searched, regardless of their language, against a query formulated in user-specific language. This is accomplished in dBlue without extra overhead or the need for a translation at runtime through a specific extension of the inverted index, a core component of most search engines. The second allows the achievement of similar results through dynamic mapping of the user's search query at run time, and use of multithreading to submit multilingual queries to the search engine. Figure 5 illustrates some aspects of this innovation.

Remote Site Customization
Another dBlue feature that addresses corporate needs is Remote Site Customization (RSC). IBM, like any other large corporation, has multiple departments that may want to present search results and technical documents to their customers in different ways, adding their own ads, promotions, and so on. The dBlue system enables this by providing the RSC feature, which allows different departments to define their own layouts for search results and technical documents. The idea of RSC is rather simple: each remote site that wants to present the shared system content in a special format is allowed to store and register its own forms. When the system gets a request that specifies this remote site, it will use the appropriate form to build the customized view of the content. Figure 6 shows the six areas that are available for customization in a results page. To assist departments in customizing the layout of Web pages, dBlue provides a Web-based RSC administrative application, which allows the uploading and testing of customized forms.

Conclusion
dBlue has many advantages. In the near future, customers will be able to ask questions in natural language and the system won't require an exact match of words. In the near future, dBlue will also personalize searching so that once a user fills out a profile, responses will be filtered and ranked based on that profile. Multilanguage searches for documents written in Japanese, Chinese, and French will be supported by late 2002. By 2Q03, it's expected that all languages will be supported from a single Web application consistent with the vision of "one Web" for all regions.

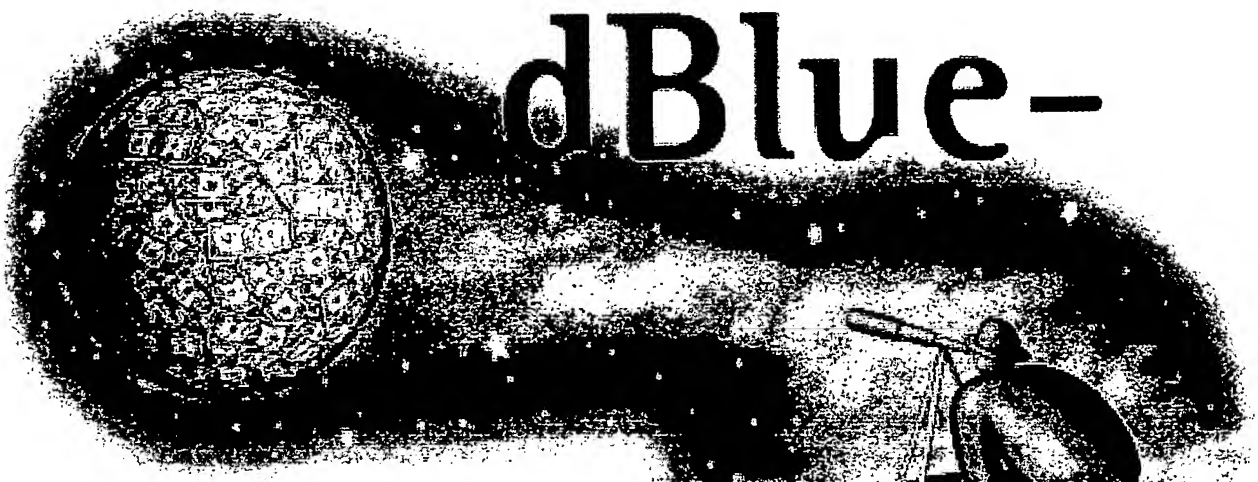
Visit Author's Web Site: Yurdaer Doganata

References

- *IBM WebSphere Software Platform Overview: www7b. boulder.ibm.com/wsdd/products/platform overview.html*
- *WebSphere Everyplace Suite: www.ibm.com/pvc/prod ucts/ wes/index.shtml*
- *Booth, Alan E. Extending the Reach of Enterprise Applications with Transcoding and Machine Translation: www7b.software.ibm.com/wsdd/library/techarti cles/0206_booth/booth.html*
- *WebSphere Technology for Developers (overview & downloads): www7b.software.ibm.com/wsdd/downloads/ wstechnology_tech_preview.html*
- *Snell James. Implementing Web services with IBM WebSphere Version 4.0: www7b.software. ibm. com/wsdd/library/techarticles/0108_snell/0108_snell. html*
- *DB2 Product Family Overview: www.ibm.com/soft ware/data/db2/*
- *About the Internationalization Activity: www.w3.org/ International/about.html*
- *Internetionalization in Jeva: http://java.sun.com/docs/ books/tutorial/i18n/*
- *The Unicode Home Page: www.unicode.org*
- *Park, Youngja; Byrd, Roy J.; and Boguraev, Branimir K. Automatic glossary extraction: Beyond terminology identification, IBM Research Technical Report RC22421, 2002. IBM T.J. Watson Research. The Talent (Text Analysis and Language Engineering) project: www.research.ibm.com/ talent*
- *Boguraev, Branimir K. and Neff, Mary S. (2000). Lexical Cohesion, Discourse Segmentation and Document Summarization. RIAO-2000. April.*
- *Park, Youngja and Byrd, Roy J. (2001). Hybrid text mining for finding terms and their abbreviations. EMNL-2001.*
- *Chu-Carrol, Jennifer; Prager, John; Ravin, Yael; and Cesar, Christian. (2002). A Hybrid Approach to Natural Language Web Services. EMNLP-2002.*

Published Oct. 21, 2002 — Reads 21,978

APPENDIX C



dBlue– An Advanced Enterprise Information Search and Delivery System

YURDAER N. DOGANATA, YOUSSEF DRISSI, TONG-HAING FIN, MOON J. KIM, AND LEV KOZAKOV

Fulfilling IBM's one Web vision

ABOUT THE AUTHOR

Dr. Yurdaer Doganata is the manager of the Integrated Networked Services group and the dBlue core research team at IBM T. J. Watson Research Center. His current focus is on alternative techniques and methods for effective information search and retrieval in unstructured and semi-structured environments. Yurdaer holds several patents and research awards and has published numerous papers. He is the technical committe cochair of ISCC'03.

E-MAIL: yurdaer@us.ibm.com

One of the biggest complaints we hear about many company Web sites, from customers and employees alike, is that it's too hard to find what you need. At IBM, with 2.5 million Internet pages and more technical content than any single entity, including the Pentagon, that's no surprise.

A new IBM advanced information search and delivery system for the IBM support site (www.ibm.com/support) is expected to solve this problem. Code-named Digital Blue (dBlue), this project is a digital interface to IBM customers. The result of two years of work and 12 patentable inventions, dBlue, is now available to IBM customers.

The team that created dBlue is calling it "the next generation of enterprise information search-and-delivery systems." This is a WebSphere-based technology with breakthroughs in storing, searching, and retrieving information. Customers will be able to search for IBM technical support information using natural language and will receive results that are categorized, prioritized, and personalized. dBlue provides the foundation for a set of user-oriented support services applicable to all IBM support sites worldwide.

Rich Vazzana, vice president of ibm.com Support and Enablement, took on this project to improve the effectiveness and performance of IBM's Web-enabled post-sales support services. It became the underlying architecture of the "one Web" vision across multiple IBM Web sites, improving adherence to IBM's company-wide standards and setting the stage for more advanced service offerings. The program will provide customers with IBM support experience, a single IBM support/service portal, a toolset, and infrastructure. Hence, cross-IBM "common" support functions will be realized.

"The business goal is to improve goal achievement on the IBM Internet," said Frank Cummiskey, director of IBM eSupport & Services. "The primary reason that customers visit IBM's support sites is to resolve a technical problem. Today, only about 60% actually achieve their goal. Improving our customers' ability to find what they are looking for, as well as to find value in the information they find, will increase self-service on the Web, saving millions of dollars and increasing customer satisfaction."

System Architecture

Although dBlue architecture does not depend on the WebSphere software platform, it's the platform of choice of the dBlue architects for its scalability, flexibility, reliability, and high performance required for dynamic Web applications hit by millions of customers every month. In addition to the application server mechanisms, the WebSphere software platform provides reliable communication middle-

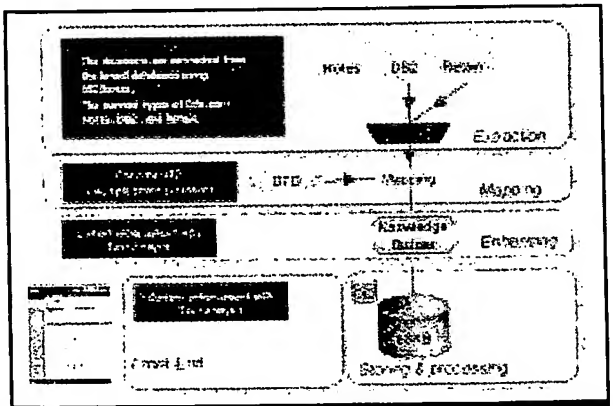


FIG. 1: DBLUE ARCHITECTURE OVERVIEW

ware - WebSphere MQ family. It also supports DB2 Universal Databases, provides a foundation for Web services, and integrates business components for text analysis and machine translation. The WebSphere Everyplace Suite provides an integrated software platform for extending the reach of business applications, enterprise data, and Internet content into the realm of pervasive computing. All this makes the WebSphere software platform the perfect fundament for the dBlue system. Figure 1 is an overview of the dBlue architecture.

The dBlue architecture connects three important elements from the information search world - information sources, search engines, and end users - on the basis of the WebSphere software platform. This is done through a set of components called "The Knowledge Builder." Information sources are data sources such as document repositories, DB2 and Lotus Notes databases, Web sites, and so on. Search engines are programs that can index content and enable searching of the indexed data. End users access dBlue through a front-end interface; the current default interface is a Web interface. The content is extracted from information sources using the Document Extractor and mapped to a unified XML Schema, then it's processed by the Document Processor and stored in the Knowledge Repository.

When a user accesses the system and submits a search query, the Query Manager, along with all the submitted parameters, processes this query. The Query Builder then collects the query and parameters submitted by the user, along with information coming from the user's profile and the system configuration, to build a standard Query object. The Query object is submitted to the search engine through the Search Engine Adapter. The search results flow back to the user through the Search Engine Adapter, the Search Query Manager, and the View Builder. The View Builder uses the Remote Site Customization component and data to construct a personalized view of the search hit list. When the user requests a view of a specific document, this request is processed by the View Builder, which accesses the Knowledge Repository to get the document content and builds a coherent document view.

Enabled by the WebSphere software platform, dBlue introduces various innovative solutions in the areas of information search and delivery. In dBlue:

- Content is indexed using the concept of virtual URLs.
- Search results and documents are rendered by employing dynamic layout features.
- Keyword and navigational search are combined for effective searching.

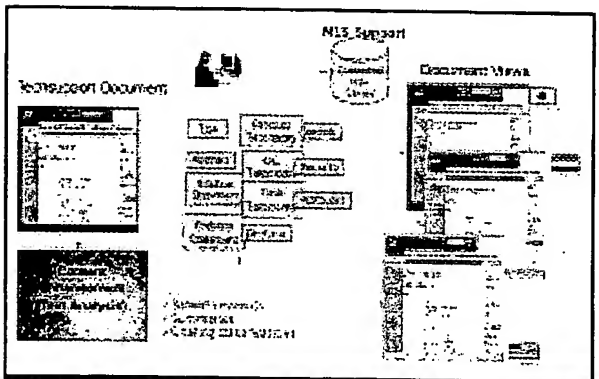


FIG. 2: CREATING MULTIPLE VIEWS FROM THE SAME CONTENT

- Search results and indexing are improved by using text analysis technologies.
- Architecture is enabled for globalization and dual language search.

Virtual URLs and Dynamic Layout

dBlue is a search system, but it doesn't depend on a particular search engine. The technical content to be indexed can be pushed to any search engine using the concept of virtual URLs. Until now, search systems have had to crawl content off a particular address where it's stored. Hence, the documents are replicated redundantly for the purpose of indexing the same information in a different context. With virtual URLs, documents to be indexed are built on-the-fly from building blocks, eliminating the need for replication and crawling. In other words, the virtual URLs aren't associated with any physically stored documents. This motivates another breakthrough in content storage. In the back end, the documents are broken down into components, such as title, problem, solution, reference, and category, allowing for true knowledge mining and the building of multiple views of the same content. Extracting the documents from their original sources and creating components based on unified XML Schema for technical documents accomplishes this, giving users a great deal of flexibility and allowing them to receive a wider range of information.

In a typical search system, the documents are stored and retrieved with a layout defined by the content providers. In this case the layout is static and cannot be changed to meet customers' needs. dBlue solves this problem by introducing the concept of dynamic layout for creating multiple views from the same content (see Figure 2).

The component-based storage system invented by the dBlue team decomposes documents into data elements without breaking the ties to their original documents. When customers request information in a specific layout, components are analyzed to ensure that they have all the necessary elements for a specific document, which is then built dynamically. This gives the flexibility to separate user experience from the content-generation process and also enables rapid localization and internationalization of the pages.

OC Taxonomy

One of the first challenges was to institute a consistent structure for content creation, since the huge amount of support content that already existed was not suitable for search. In order to structure the content and organize the content-creation process, the unified XML Schema for technical documents was created. This schema incorporates content components, such as title, abstract, problem statement, and solution statement, along with multiple attributes, keywords, references, and attachments.

The second step in organizing the content was creation of the content repository schema that allows storage of both unstructured and structured data. This schema contains more than 30 DB2 tables that provide storage for the document content, along with all associated information, and supports a variety of queries. Then, of course, both existing and new content had to be migrated to this structure. The content migration pipe is powered by the WebSphere MQ family communication middleware. The documents extracted from their original repositories were converted to XML format based on the unified XML Schema and transferred to the new storage. All document attachments were encoded

ABOUT THE AUTHOR

Youssef Drissi is an advisory software engineer at the IBM T.J. Watson Research Center in Hawthorne, New York. He holds several patents and is the author of several publications. Youssef is a member of the dBlue project architecture and research teams. His current work involves research, architecture, and development of next-generation unstructured information and knowledge management systems.

E-MAIL: youssefd@us.ibm.com

ABOUT THE AUTHOR

Dr. Tong Fin is a senior software engineer at the IBM T.J. Watson Research Center. He holds a number of patents and is the author of several publications. He is a member of the dBlue project architecture and research teams. His current work involves research and system integration of text analysis and knowledge management systems.

E-MAIL: tfin@us.ibm.com

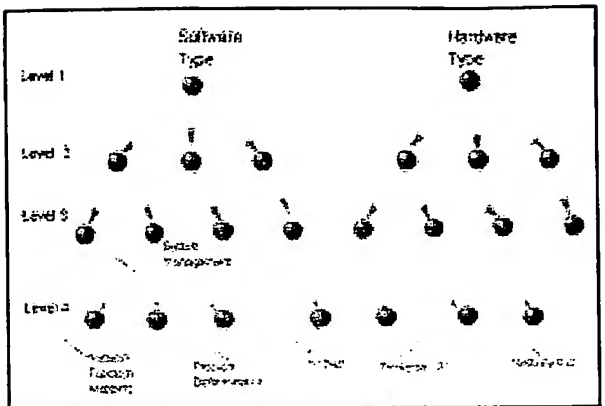


FIG. 3: OC TAXONOMY TREE FRAGMENTS

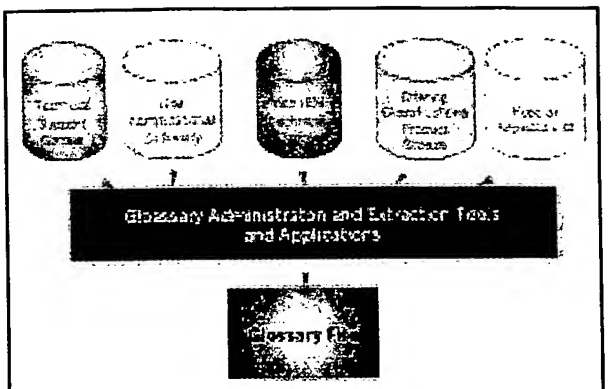


FIG. 4: GLOSSARY OF TECHNICAL TERMS

using "Base64" encoding and incorporated in XML objects. To eliminate unnecessary XML parsing, the transportation was done in a binary format.

Another challenge was determining how to store and dynamically retrieve this information in a scalable and flexible way. The team adopted a categorization scheme based on IBM product offerings, called offering classification (OC) in IBM. The common library classification can be used, but for the IBM technical support all contents are associated with IBM products. With the OC taxonomy attached to the content, the content can easily be shown where it belongs. Figure 3 shows a fragment of the OC taxonomy tree with sample documents that may be found under certain leaves.

Having OC taxonomy information attached to the documents made it possible to combine a keyword with the navigational search. This way, users can narrow down search results with single click.

Combining Keyword with Navigation Search

The way the system is architected allows combining keyword search with navigational search. Based on a topic or a document type, users can narrow down search findings with a single click. This increases the chances of finding the requested information when the user query isn't specific enough to narrow down the search results on the first attempt. The categorized results are returned with the option of filtering the results based on IBM's product offerings and the document types.

Although combining keyword and navigational search helps refine the search results, it doesn't improve relevancy or precision/recall rates. The following sections discuss some text-analysis techniques used to improve precision/recall.

Content Enhancement for Search Improvement

The quality of full text search depends mainly on query terms and on how documents are indexed by the search engine. The search results contain the documents that are indexed against the query terms and scored based on certain statistical criteria. In many real-life situations, the relevant documents can't be found or may not appear at the top of the search results because they are scored low or they don't contain the terms exactly as in the query. This is common when users choose variations of the query terms, including inflections, misspellings, abbreviations, and so on. To improve the user experience, dBlue uses text analysis tools developed by IBM Research to enhance the contents of documents. This process is started by extracting terms from a large collection of documents in the IBM technical support domain to create a domain-specific glossary. The terms in the glossary can consist of canonical form, variant form (inflection, abbreviation, misspelling, etc.), synonym, term definition, statistical data, and other information. This initial glossary is enhanced by eliminating irrelevant terms and reranking terms using special dictionaries and algorithms. The process of generating and enhancing the glossary is semi-automatic, using glossary tools and the librarian. Figure 4 shows multiple components that comprise the glossary of technical terms built for the dBlue system.

Based on the glossary, the important keywords in each document are extracted and ranked, and their related glossary terms (variants, synonyms, etc.) are used to enrich the content of the document. The content enrichment is used to create keyword metatags for biased indexing, expand the query terms to include related terms, and enable search for related documents. To improve the user's search experience, keywords are displayed in the search results and navigating through keywords is possible.

Globalization

As part of the effort to allow different languages to be supported from a single Web application consistent with the vision of "one Web" for all regions, dBlue was enabled with a globalization process that consists of two main processes: internationalization and localization.

INTERNALIZATION

Internationalization (sometimes abbreviated as i18n) is the process of designing an application so that it can be adapted to various languages and regions without engineering changes. After the internalization of dBlue software components, they can run worldwide with the addition of localized data. Hence, support for new languages doesn't require recompilation. Textual elements, such as status messages and the GUI component labels, are stored outside of the source code and retrieved dynamically. Culturally dependent data, such as dates and currencies, appears in formats that conform to the end user's region and language.

The Unicode format, which handles most characters known to mankind, was instrumental in allowing the use of a unique globalized repository where multi-lingual

ABOUT THE AUTHOR

Moon J. Kim is an IBM senior technical staff member with responsibility for the development of the e-Support advanced Web system. Moon also developed many large system solutions and was involved in the development of the network systems that later called the broadband high-speed access system including HFC and FSN. Moon is an IBM Master Inventor and holds 10 patents, has published 10 invention technical papers, and has filed 10.

E-MAIL: moonkim@us.ibm.com

searchable text and documents are encoded in one unique format. Unicode was also adopted as a standard format for encoding internal textual data in dBlue.

LOCALIZATION

Localization (sometimes abbreviated as i10n) is the process of adapting software for a specific region or language by adding locale-specific components and translating text. Usually, the most time-consuming part of the localization process is the translation of text. Other types of data, such as sounds and images, may require localization if they are culturally sensitive. Localizers also verify that the formatting of dates, numbers, and currencies conforms to local requirements.

Two innovative approaches in the globalization process are worth mentioning. The first allows documents to be searched, regardless of their language, against a query formulated in user-specific language. This is accomplished in dBlue without extra overhead or the need for a translation at runtime through a specific extension of the inverted index, a core component of most search engines. The second allows the achievement of similar results through dynamic mapping of the user's search query at runtime, and use of multithreading to submit multilingual queries to the search engine. Figure 5 illustrates some aspects of this innovation.

Remote Site Customization

Another dBlue feature that addresses corporate needs is Remote Site Customization (RSC). IBM, like any other large corporation, has multiple departments that may want to present search results and technical documents to their customers in different ways, adding their own ads, promotions, and so on. The dBlue system enables this by providing the RSC feature, which allows different departments to define their own layouts for search results and technical documents. The idea of RSC is rather simple: each remote site that wants to present the shared system content in a special format is allowed to store and register its own forms. When the system gets a request that specifies this remote site, it will use the appropriate form to build the customized view of the content. Figure 6 shows the six areas that are available for customization in a results page. To assist departments in customizing the layout of Web pages, dBlue provides a Web-based RSC administrative application, which allows the uploading and testing of customized forms.

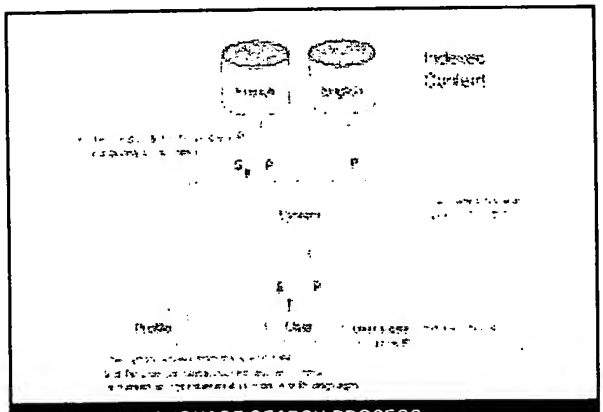


FIG. 5: MULTILANGUAGE SEARCH PROCESS

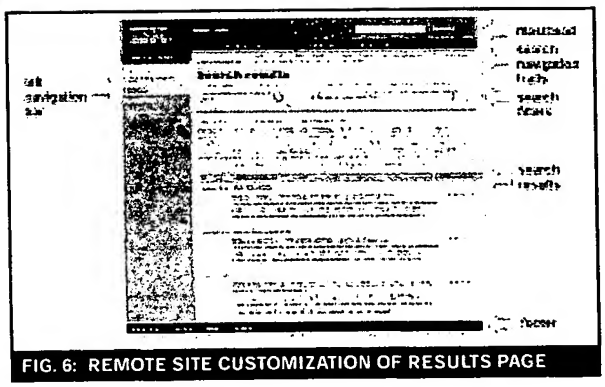


FIG. 6: REMOTE SITE CUSTOMIZATION OF RESULTS PAGE

Conclusion

dBlue has many advantages. Customers will be able to ask questions in natural language and the system won't require an exact match of words. In the near future, dBlue will also personalize searching so that once a user fills out a profile, responses will be filtered and ranked based on that profile. Multilanguage searches for documents written in Japanese, Chinese, and French will be supported in the next version later this year. By 2Q03, it's expected that all languages will be supported from a single Web application consistent with the "one Web" for all regions vision.

References

- *IBM WebSphere Software Platform Overview:* www7b.boulder.ibm.com/wsdd/products/platformoverview.html
- *WebSphere Everyplace Suite:* www.ibm.com/pvc/products/wes/index.shtml
- Booth, Alan E. *Extending the Reach of Enterprise Applications with Transcoding and Machine Translation:* www7b.software.ibm.com/wsdd/library/techarticles/0206_booth/booth.html
- *WebSphere Technology for Developers (overview & downloads):* www7b.software.ibm.com/wsdd/downloads/wstechnology_tech_preview.html
- Snell James. *Implementing Web services with IBM WebSphere Version 4.0:* http://www7b.software.ibm.com/wsdd/library/techarticles/0108_snell/0108_snell.html
- *DB2 Product Family Overview:* www.ibm.com/software/data/db2/
- *About the Internationalization Activity:* www.w3.org/International/about.html
- *Internationalization in Java:* http://java.sun.com/docs/books/tutorial/i18n/
- *The Unicode Home Page:* www.unicode.org
- Park, Youngja; Byrd, Roy J; and Boguraev, Branimir K. *Automatic glossary extraction: Beyond terminology identification, IBM Research Technical Report RC22421, 2002. IBM T.J. Watson Research. The Talent (Text Analysis and Language Engineering) project:* www.research.ibm.com/talent/.
- Boguraev, Branimir K. and Neff, Mary S. (2000). *Lexical Cohesion, Discourse Segmentation and Document Summarization.* RIAO-2000. April.
- Park, Youngja and Byrd, Roy J. (2001). *Hybrid text mining for finding terms and their abbreviations.* EMNL-2001.
- Chu-Carrol, Jennifer; Prager, John; Ravin, Yael; and Cesar, Christian. (2002). *A Hybrid Approach to Natural Language Web Services.* EMNLP-2002.

ABOUT THE AUTHOR

Dr. Lev Kozakov is a research staff member at IBM T.J. Watson Research Center and is a member of the dBlue project architecture and research teams. He has worked in many areas, including dynamic systems, applied statistics, information management systems, man-machine interface, medical software, computer telephony, and design patterns. Lev holds a number of patents and is the author of several publications.

E-MAIL: kozakov@us.ibm.com

APPENDIX D

Websphere Authors: Kym Jones, Theresa Lanowitz, Reuven Cohen, Maureen O'Gara, Jeremy Geelan
Related Stories: Websphere

Websphere: Article

dBlue - An Advanced Enterprise Information Search and Delivery System

*The result of two years of work and five patentable inventions, **dBlue** is now available to IBM customers*

BY YURDAER DOGANATA, YOUSSEF DRISSI, LEV KOZAKOV, TONG-HAING FIN, GREG BROWN, MOON J. KIM
OCTOBER 21, 2002 05:45 PM EDT

ARTICLE RATING:

READS: 22,834

(Article Publish Date: October 21, 2002) - One of the biggest complaints we hear about many company Web sites, from customers and employees alike, is that it's too hard to find what you need. At IBM, with 2.5 million Internet pages and more technical content than any single entity, including the Pentagon, that's no surprise.

A new IBM advanced information search and delivery system for the IBM support site (www.ibm.com/support) is expected to solve this problem. Code-named Digital Blue (**dBlue**), this project is a digital interface to IBM customers. The result of two years of work and five patentable inventions. **dBlue** is now available to IBM customers.

The team that created **dBlue** is calling it "the next generation of enterprise information search-and-delivery systems." This is a WebSphere-based technology with breakthroughs in storing, searching, and retrieving information. Customers will be able to search for IBM technical support information using natural language and will receive results that are categorized, prioritized, and personalized. **dBlue** provides the foundation for a set of user-oriented support services applicable to all IBM support sites worldwide.

Rich Vazzana, vice president of ibm.com Support and Enablement, took on this project to improve the effectiveness and performance of IBM's Web-enabled post-sales support services. It became the underlying architecture of the "one-Web" vision across multiple IBM Web sites, improving adherence to IBM's company-wide standards and setting the stage for more advanced service offerings. The program will provide customers with IBM support experience, a single IBM support/service portal, toolset, and infrastructure. Hence, cross-IBM "common" support functions will be realized.

"The business goal is to improve goal achievement on the IBM Internet," said Frank Cummiskey, director of IBM eSupport & Services. "The primary reason that customers visit IBM's support sites is to resolve a technical problem. Today, only about 60% actually achieve their goal. Improving our customers' ability to find what they are looking for, as well as to find value in the information they find, will increase self-service on the Web, saving millions of dollars and increasing customer satisfaction."

System Architecture

Comments

JavaFX vs Silverlight vs Flex

Lauryn Oates wrote: I need one of them for a integration with Sharepoint, I tried Sharepoint Java solution by Mainsoft and seems like Flex is no good. Could JavaFX work with Mainsoft's solution?

Mar. 29, 2009 05:42 AM EDT

Topics

All Sections
WSDJ Archives
Application Management
Application Server
Automation
Business Process Modeling
Case Study

customized forms.

Conclusion

dBlue has many advantages. In the near future, customers will be able to ask questions in natural language and the system won't require an exact match of words. In the near future, **dBlue** will also personalize searching so that once a user fills out a profile, responses will be filtered and ranked based on that profile. Multilanguage searches for documents written in Japanese, Chinese, and French will be supported by late 2002. By 2Q03, it's expected that all languages will be supported from a single Web application consistent with the vision of "one Web" for all regions.

Visit Author's Web Site: Yurdaer Doganata

References

- *IBM WebSphere Software Platform Overview: www7b. boulder.ibm.com/wsdd/products/platform overview.html*
- *WebSphere Everyplace Suite: www.ibm.com/pvc/prod ucts/ wes/index.shtml*
- *Booth. Alan E. Extending the Reach of Enterprise Applications with Transcoding and Machine Translation: www7b.software.ibm.com/wsdd/library/techarti cles/0206_booth/booth.html*
- *WebSphere Technology for Developers (overview & downloads): www7b.software.ibm.com/wsdd/downloads/ wstechnology_tech_preview.html*
- *Snell James. Implementing Web services with IBM WebSphere Version 4.0: www7b.software. ibm. com/wsdd/library/techarticles/0108_snell/0108_snell. html*
- *DB2 Product Family Overview: www.ibm.com/soft ware/data/db2/*
- *About the Internationalization Activity: www.w3.org/ International/about.html*
- *Internationalization in Java: http://java.sun.com/docs/ books/tutorial/i18n/*
- *The Unicode Home Page: www.unicode.org*
- *Park, Youngja; Byrd, Roy J.; and Boguraev, Branimir K. Automatic glossary extraction: Beyond terminology identification, IBM Research Technical Report RC22421, 2002. IBM T.J. Watson Research. The Talent (Text Analysis and Language Engineering) project: www.research.ibm.com/ talent*
- *Boguraev, Branimir K. and Neff, Mary S. (2000). Lexical Cohesion, Discourse Segmentation and Document Summarization. RIAO-2000. April.*
- *Park, Youngja and Byrd, Roy J. (2001). Hybrid text mining for finding terms and their abbreviations. EMNL-2001.*
- *Chu-Carrol, Jennifer; Prager, John; Ravin, Yael; and Cesar, Christian. (2002). A Hybrid Approach to Natural Language Web Services. EMNLP-2002.*

Published October 21, 2002 – Reads 22,834

Related Links

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

About Yurdaer Doganata

Dr. Doganata is the manager of the Information Management Solutions group at the Watson Research Center in Hawthorne, New York. He received B.S. and M.S. degrees from the Middle East Technical University, Ankara, Turkey, and a Ph.D. degree from the California Institute of Technology, Pasadena, California, all in electrical engineering. He joined the Watson Research Center as a research staff member in 1989 and worked on projects in many diverse areas, including high-speed switching systems, multimedia servers, intelligent transportation systems, multimedia collaborative applications, eservices, and